In the claims:

1. (currently amended) A computer-implemented method of identifying table data in a document comprising the steps of:

a) receiving a page description language representation of the document for providing a list of words in the document and position information for the words; and

b) automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein the identifying step includes,

b1) dividing the document into one or more pages;

b2) dividing each page into a plurality of lines;

b3) for each line, clustering the words of the line into one or more word clusters;

b4) automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters between lines.

2. (canceled)

3. (currently amended) The method of Claim 1 ~~2~~ wherein the step of automatically identifying table data in the document based on the number of word clusters of each line and the alignment of the word clusters between lines further comprises:

b4_1) using the word clusters to generate column position information; and

b4_2) updating the column position information by performing a union operation between the column position information of the previous line and the column position information of the current line.

4. (currently amended) ~~The method of Claim 1~~ A computer-implemented method of identifying table data in a document comprising the steps of:

a) receiving a page description language representation of the document for providing a list of words in the document and position information for the words; and

b) automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein said step of automatically identifying table data in the document based on the page description of the document and at least one table identifying feature includes, ~~comprises:~~

b1) automatically determining a table bounding box for each table in the document;

b2) expanding each table bounding box based on a text density feature; and

b3) converting the table data encompassed by each table bounding box to a markup language representation.

5. (original) The method of Claim 4 wherein receiving a page description language representation of the document for providing a list of words in the document and position information for the words includes receiving a PDF representation of the document, and wherein converting the table data encompassed by each table bounding box to a markup language representation includes converting the table data encompassed by each table bounding box to a HTML representation.

6. (canceled)

7. (currently amended) A computer-readable medium having stored thereon sequences of instructions, said sequences of instructions including instructions which, when executed by a processor, cause said processor to perform the steps of:

a) receiving a page description language representation of a document for providing a list of words in the document and position information for the words; and

b) automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein identifying includes,

b1) dividing the document into one or more pages;

b2) dividing each page into a plurality of lines;

b3) for each line, clustering the words of the line into one or more word clusters; and

b4) automatically identifying table data in the document based on the number of word clusters for each line and the alignment of the word clusters between lines.

8. (canceled)

9. (currently amended) The computer-readable medium of Claim 7 ~~8~~ further containing instructions which, when executed by said processor, would cause said processor to perform the steps of:

b4_1) using the word clusters to generate column position information; and

b4_2) updating the column position information by performing a union operation between the column position information of the previous line and the column position information of the current line.

10. (currently amended) ~~The computer-readable medium of Claim 7 further containing instructions which, when executed by said processor, would cause said processor to perform the steps of:~~

A computer-readable medium having stored thereon sequences of instructions, said sequences of instructions including instructions which, when executed by a processor, cause said processor to perform the steps of:

a) receiving a page description language representation of a document for providing a list of words in the document and position information for the words; and

b) automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature, wherein identifying includes,

b1) automatically determining a table bounding box for each table in the document;

b2) expanding each table bounding box based on a text density feature; and

b3) converting the table data encompassed by each table bounding box to a markup language representation.

11. (canceled)

12. (currently amended) A document processing system comprising:

a) a processor for executing programs; and

b) a table identification program for receiving a page description language representation of a document, the page description language representation providing a list of words in the document and position information for the words, and for automatically identifying table data in the document based on the page description representation of the document and at least one table identifying feature, wherein the identification program includes a bounding box generation module for receiving the list or words and for automatically generating a table bounding box for each table in the document based on the number of work clusters in each line.

13. (canceled)

14. (canceled)

15. (currently amended) The document processing system of claim 12 ~~13~~ wherein the table identification program further comprises:

b3) a conversion module coupled to the bounding box generation module for receiving the table bounding box for each table in the document, and for converting the words encompassed by the table bounding box into a markup language representation that maintains the table structure of each table.

16. (original) The method of claim 1 wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

b1) automatically identifying table data in the document based on one or more table headings.

17. (original) The method of claim 1 wherein the step of automatically identifying table data in the document based on the page description language representation of the document and at least one table identifying feature further comprises:

b1) automatically identifying table data in the document based on one or more horizontal lines and vertical lines that separate rows or columns of the table.